An Evidential Improvement for Gender Profiling

Jianbing Ma, Weiru Liu, and Paul Miller

School of Electronics, Electrical Engineering and Computer Science,
Queen's University Belfast, Belfast BT7 1NN, UK
{jma03,w.liu}@qub.ac.uk, p.miller@ecit.qub.ac.uk

Abstract. CCTV systems are broadly deployed in the present world. To ensure in-time reaction for intelligent surveillance, it is a fundamental task for real-world applications to determine the gender of people of interest. However, normal video algorithms for gender profiling (usually face profiling) have three drawbacks. First, the profiling result is always uncertain. Second, for a time-lasting gender profiling algorithm, the result is not stable. The degree of certainty usually varies, sometimes even to the extent that a male is classified as a female, and vice versa. Third, for a robust profiling result in cases were a person's face is not visible, other features, such as body shape, are required. These algorithms may provide different recognition results - at the very least, they will provide different degrees of certainties. To overcome these problems, in this paper, we introduce an evidential approach that makes use of profiling results from multiple algorithms over a period of time. Experiments show that this approach does provide better results than single profiling results and classic fusion results.

Keywords: Gender Profiling; Gender Recognition; Evidence Theory; Dempster-Shafer Theory; Cautious Conjunctive Rule; Dempster's Rule.

1 Introduction

During the last decade, there has been massive investment in CCTV technology in the UK, e.g., e.g., the First Glasgow Bus Surveillance [10], Intelligent Surveillance Project [3–8], Airport Corridor Surveillance [9], etc. Currently, there are approximately four million CCTV cameras operationally deployed. Despite this, the impact on anti-social and criminal behaviour has been minimal. For example, assaults on bus and train passengers are still a major problem for transport operators. Although most incidents, also called events, are captured on video, there is no response because very little of the data is actively analyzed in real-time. Consequently, CCTV operates in a passive mode, simply collecting enormous volumes of video data. For this technology to be effective, CCTV has to become active by alerting security analysts in real-time so that they can stop or prevent the undesirable behaviour. Such a quantum leap in capability will greatly increase the likelihood of offenders being caught, a major factor in crime prevention.

A key requirement for active CCTV systems is to automatically determine the threat posed by each individual to others in the scene. Most of the focus of the computer vision community has been on behaviour/action recognition. However, experienced security analysts profile individuals in the scene to determine their threat. Often they can identify individuals who look as though they may cause trouble before any anti-social behaviour has occurred. From criminology studies, the vast majority of offenders are young adolescent males. Therefore, key to automatic threat assessment is to be able to automatically profile people in the scene based on their gender and age. In this paper, we focus on the former.

The most obvious cue in determining a person's gender is the appearance of their face. However, for automatic classifiers this usually requires cooperative subjects who are directly looking at the camera and at close range. For most security scenarios one cannot assume this, as the person's face may not be visible as they are facing away from the camera, or they may be too far away - the resulting low resolution making gender discrimination difficult or impossible. Another obvious cue that can help overcome these issues is that of body shape. However, generally automatic classifiers of body shape are a less reliable indicator of gender than face-based classifiers. Furthermore, for both types of classifiers, the output result always has some degree of uncertainty. Secondly, when such classifiers are applied to video sequences, their output can vary significantly with time - even to the extent that a person's gender is incorrectly classified. Thirdly, the key to a robust solution is to use both face and body shape classifiers. Ideally, we would like to use the face classifier result, provided it is detected, otherwise we should resort to using the body shape result. However, this raises the issue of what to do when the outputs of both classifiers are different.

Imperfect information frequently occurs in video analytic processes. For example, a person may be classified as male with a certainty of 85% by a gender profiling algorithm. However, this does not imply that the person is female with a 15% certainty, rather, we say that the 15% represents what is unknown about the gender, i.e., we do not know how to distribute the remaining 15% between male and female. From probability theory, this information can only be represented as $p(male) \geq 0.85$ and $p(female) \leq 0.15$ (or interval probabilities), which is difficult to use for reasoning. Imperfect information is usually caused by ignorance or unreliability of the information sources. For example, a camera may have a faulty gain control setting, illumination could be poor, or the classifier training set may be unrepresentative. Any, or all, of these can result in imperfect information which cannot be represented by probability measures. On the other hand, such imperfect information can be easily handled using an evidential approach, namely, the Dempster-Shafer (DS) theory of evidence.

To address all of the above issues, we investigate whether a DS framework can combine uncertain profiling results from face and body shape classifiers over an extended time period, to provide robust gender profiling of subjects in video. Experiments show that this approach provides better results than a probabilistic approach. DS theory [1, 11] is a popular framework to deal with uncertain or incomplete information from multiple sources. This theory is capable of modelling incomplete information through ignorance. For combining difference pieces of information, DS theory distinguishes two cases, i.e., whether pieces of information are from distinct, or non-distinct, sources. Many combination rules are proposed for information from distinct sources, among which are the well-known Dempster's rule [11] and Smets' rule [12]. In [2], two combination rules, i.e., the cautious rule and the bold disjunctive rule, for information from non-distinct sources are proposed. Thus, we view gender profiling results from the same classifier, e.g. face-based, at different times as being from non-distinct sources. For profiling results from different classifiers, they are naturally considered as being from distinct sources. Therefore, all of the problems mentioned above can be handled within the DS framework.

To the best of our knowledge, our approach is the first that addresses imperfect information from multiple sources for gender profiling. We demonstrate the significance and usefulness of our framework with experimental results on sample videos and by comparison to a probabilistic approach.

The rest of the paper is organized as follows. Section 2 provides the preliminaries on Dempster-Shafer theory. In Section 3, we discuss the difficulties in gender profiling in terms of scenarios. Section 4 provides experimental results which shows our method is better than a classic fusion approach and single profiling approaches. Finally, we conclude the paper in Section 5.

2 Dempster-Shafer Theory

For convenience, we recall some basic concepts of Dempster-Shafer's theory of evidence. Let Ω be a finite, non-empty set called the frame of discernment, denoted as, $\Omega = \{w_1, \cdots, w_n\}$.

Definition 1 *A basic belief assignment(bba) is a mapping* $m : 2^\Omega \rightarrow [0, 1]$ *such that* $\sum_{A \subseteq \Omega} m(A) = 1$.

If $m(\emptyset) = 0$, then m is called a mass function. If $m(A) > 0$, then A is called a focal element of m. Let $\mathscr{F}_m$ denote the set of focal elements of m. A mass function with only a focal element Ω is called a *vacuous* mass function.

From a bba m, *belief* function (Bel) and *plausibility* function (Pl) can be defined to represent the lower and upper bounds of the beliefs implied by m as follows.

$$Bel(A) = \sum_{B \subseteq A} m(B) \text{ and } Pl(A) = \sum_{C \cap A \neq \emptyset} m(C). \tag{1}$$

One advantage of DS theory is that it has the ability to accumulate and combine evidence from multiple sources by using *Dempster's rule of combination*. Let m_1 and m_2 be two mass functions from two distinct sources over Ω. Combining m_1 and m_2 gives a new mass function m as follows:

$$m(C) = (m_1 \oplus m_2)(C) = \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A) m_2(B)} \tag{2}$$

In practice, sources may not be completely reliable, to reflect this, in [11], a *discount rate* was introduced by which the mass function may be discounted in order to reflect the reliability of a source. Let r $(0 \leq r \leq 1)$ be a discount rate, a discounted mass function using r is represented as:

$$m^r(A) = \begin{cases} (1-r)m(A) & A \subset \Omega \\ r + (1-r)m(\Omega) & A = \Omega \end{cases} \tag{3}$$

When $r = 0$ the source is absolutely reliable and when $r = 1$ the source is completely unreliable. After discounting, the source is treated as totally reliable.

Definition 2 *Let m be a bba on Ω. A pignistic transformation of m is a probability distribution P_m over Ω such that $\forall w \in \Omega, P_m(w) = \sum_{w \in A} \frac{1}{|A|} \frac{m(A)}{1-m(\emptyset)}$ where $|A|$ is the cardinality of A.*

Let $\oplus$ be the conjunctive combination operator (or Smets' operator [12]) for any two bbas m, m' over Ω such that

$$(m \oplus m')(C) = \sum_{A \subseteq \Omega, B \subseteq \Omega, A \cap B = C} m(A)m'(B), \forall\, C \subseteq \Omega. \tag{4}$$

A simple bba m such that $m(A) = x, m(\Omega) = 1 - x$ for some $A \neq \Omega$ will be denoted as A^x. The vacuous bba can thus be noted as A^0 for any $A \subset \Omega$. Note that this notation, i.e., A^x, is a bit different from the one defined in [2] in which A^x in our paper should be denoted as A^{1-x} in [2].

Similarly, for two sets $A, B \subset \Omega$, $A \neq B$, let $A^x B^y$ denote a bba m such that $m = A^x \oplus B^y$ where $\oplus$ is the conjunctive combination operator defined in Equation (4). For these kinds of bbas, we call them *bipolar* bbas. A simple bba A^x could be seen as a special bipolar bba $A^x B^0$ for any set $B \subseteq \Omega, B \neq A$.

It is easy to prove that any $m = A^x B^y$ is:

$$m(\emptyset) = xy, m(A) = x(1-y), m(B) = y(1-x), m(\Omega) = (1-x)(1-y) \tag{5}$$

In addition, when normalized, m in Equation 5 is changed to m' as follows.

$$m'(A) = \frac{x(1-y)}{1-xy}, m'(B) = \frac{y(1-x)}{1-xy}, m'(\Omega) = \frac{(1-x)(1-y)}{1-xy} \tag{6}$$

For two bipolar bbas $A^{x_1}B^{y_1}$ and $A^{x_2}B^{y_2}$, the cautious combination rule proposed in [2] is as follows.

Lemma 1 *(Denœux's Cautious Combination Rule) Let $A^{x_1}B^{y_1}$ and $A^{x_2}B^{y_2}$ be two bipolar bbas, then the combined bba by Denœux's cautious combination rule is also a bipolar bba $A^x B^y$ such that: $x = max(x_1, x_2), y = max(y_1, y_2)$.*

Also, according to [2], for $m_1 = A^{x_1}B^{y_1}$ and $m_2 = A^{x_2}B^{y_2}$, the combined result by Equation (2) is

$$m_{12} = A^{x_1 x_2} B^{y_1 y_2} \tag{7}$$

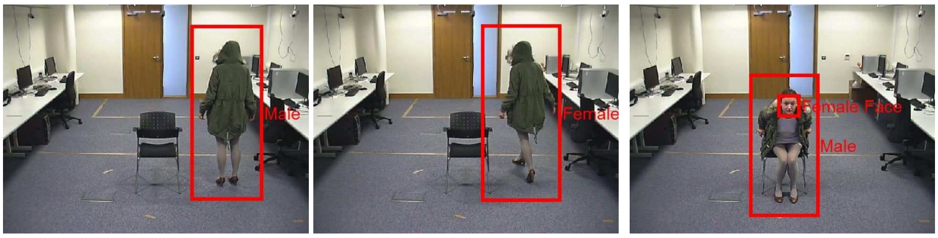
3 Gender Recognition Scenario

In this section, we provide a detailed description of a gender profiling scenario, which lends itself naturally to a DS approach.

Figure 1 shows three images taken from a video sequence that has been passed through a video analytic algorithm for gender profiling. In this sequence, a female wearing an overcoat with a hood enters the scene with her back to the camera. She walks around the chair, turning, so that her face becomes visible, and then sits down.

Fig. 1(a) shows that the subject is recognised by the full body shape profiling as a male. Note that her face is not visible. In Fig. 1(b), the subject is classified as female by

Fig. 1. Three images taken from a video sequence



(a) (b) (c)

the full body shape profiling algorithm. In Fig. 1(c), as she sits down, with her face visible, the face profiling algorithm classifies her as female, whilst the full body profiling classifies her as male. Note that the full body profiling algorithm is not as reliable as the face profiling algorithm. Conversely, full body profiling is always possible whilst the face information can be missing. That is why these two profiling algorithms should be considered together. In addition, as full body profiling is not as robust, discount operations should be performed on the algorithm output (cf. Equation (3)). The discount rate is dependent on the video samples and the training efficiency. For every video frame in which a body (face) is detected, gender recognition results are provided. The full body profiling algorithm and the face profiling algorithm, provided a person's face is detected, report their recognition results for every frame of the video, e.g., male with 95% certainty.

For a frame with only a body profiling result, for instance Fig. 1(a), the corresponding mass function m for body profiling will be M^x where M denotes that *the person is classified as a male* and x is the mass value of $m(\{M\})$. The corresponding mass function for face profiling is M^0F^0 where F denotes that *the person is classified as a female*, or the vacuous mass function. Alternatively, we can refer to this as the vacuous mass function.

Similarly, for a frame with both body profiling and face profiling, for instance Fig. 1(c), the corresponding mass function for body profiling will be M^x (or in a bipolar form M^xF^0) and the mass function for face profiling is F^y (or in a bipolar form M^0F^y) where x, y are the corresponding mass values. As time elapses, fusion of bipolar bbas by the cautious rule is reduced, as shown by Lemma 1. And when it comes to present the final profiling result, we use Dempster's rule to combine the two fused bipolar mass functions from the two recognition algorithms, respectively. Namely, for the two bipolar bbas $m_1 = M^{x_1}F^{y_1}$ and $m_2 = M^{x_2}F^{y_2}$, it is easy to get that the combined result m_{12} by Dempster's rule is (normalized from the result of Equation 7):

$$m_{12}(\{M\}) = \frac{m_1(\{M\}m_2(\{M\}))(1 - m_1(\{F\})m_2(\{F\}))}{1 - m_1(\{M\})m_2(\{M\})m_1(\{F\})m_2(\{F\})},$$

$$m_{12}(\{F\}) = \frac{m_1(\{F\}m_2(\{F\}))(1 - m_1(\{M\})m_2(\{M\}))}{1 - m_1(\{M\})m_2(\{M\})m_1(\{F\})m_2(\{F\})},$$

$$m_{12}(\Omega) = \frac{(1 - m_1(\{M\})m_2(\{M\}))(1 - m_1(\{F\})m_2(\{F\}))}{1 - m_1(\{M\})m_2(\{M\})m_1(\{F\})m_2(\{F\})}.$$

Finally, we use the pignistic transformation (Def. 2) for the final probabilities. That is, $p(\{M\}) = m_{12}(\{M\}) + m_{12}(\Omega)/2$ and $p(\{F\}) = m_{12}(\{F\}) + m_{12}(\Omega)/2$

Example 1 *Let us illustrate the approach by a simple scenario with two frames. In the first frame, we have both body profiling (m_b^1) and face profiling (m_f^1) results as $m_b^1 = M^{0.7}F^{0.3}$ and $m_f^1 = M^{0.4}F^{0.6}$. In the second frame, we have the body profiling (m_b^2) result only, where $m_b^2 = M^0.8F^0.2$. By Lemma 1, the fusion results by the cautious rule is $m_b = M^{0.8}F^{0.3}$ and $m_f = M^{0.4}F^{0.6}$. Then by Equation 7, we get $m_{bf} = M^{0.32}F^{0.18}$, which, when normalized, is equivalent to $m_{bf}(\{M\}) = \frac{0.32(1-0.18)}{1-0.32*0.18} = 0.28$, $m_{bf}(\{F\}) = \frac{0.18(1-0.32)}{1-0.32*0.18} = 0.13$, $m_{bf}(\Omega) = \frac{(1-0.32)(1-0.18)}{1-0.32*0.18} = 0.59$. And finally we get $p(\{M\}) = 0.58$ and $p(\{F\}) = 0.42$.*

4 Experimental Results

In this section we compare fusion results obtained by Dempster-Shafer theory and a classic approach. As there are no benchmark datasets for both body and face profiling, we simulate the output of both body and face classifiers on a sequence containing a male subject. For the body classifier, the probability of any frame being correctly classified as male/female is roughly 60-90%. For the face classifier, only 75% of the available frames are randomly allocated as containing a face. For each of these frames the probability of the frame being correctly classified as being male/female is 85-100%. In both cases the values for $m(\{M\})$ and $m(\{F\})$ are uniformly sampled from the ranges 0.6-0.9 and 0.85-1.0 for the body and face classifiers outputs respectively.

As mentioned before, for gender profiling results from the same classifier at different time points, we use the cautious rule (Lemma 1) to combine them. For profiling results from different classifiers (i.e., face profiling and full body profiling), we use Dempster's rule (Equation (2)) to combine them. And finally, we apply the pignistic transformation (Def. 2) to get the probabilities of the subject being male or female.

Classic fusion in the computer vision community [13] takes the degrees of certainty as probabilities, i.e., they consider the face profiling and the full body profiling output p_f^t and p_b^t indicating the probabilities of faces and full bodies being recognized as males at time t. Then it uses $p_{b,f}^t = c_f^t p_f^t + c_b^t p_b^t$ to calculate the final probability $p_{b,f}^t$ at time t, where c_f^t and c_b^t are the weights of the face and full body profiling at time t, proportional to the feasibility of the two algorithms in the last twenty frames. As full body profiling is always feasible, suppose face profiling can be applied n times in the last twenty frames, then we have:

$$c_b = \frac{20}{20+n}, c_f = \frac{n}{20+n}.$$

For this experiment, the performance of the DS and classic fusion schemes were characterised by the true positive rate:

$$T_{PR} = \frac{N_{PR}}{N}$$

where N_{PR} is the number of frames in which the gender has been correctly classified and N is the total number of frames in which the body/face is present. According to the training on the sample videos, the discount rate r for the full body profiling is set to 0.3. For comparison, we calculate the T_{PR} value for the body classifier alone, the face classifier, the DS fusion scheme and the classic fusion scheme.

When applying the methods on the randomly-generated simulation data, the comparison results are presented as follows.

Methods	TotalFrame	N	N_{PR}	T_{PR} (%)
Full Body	3100	3100	1872	60.4
Face	3100	2321	2178	93.8
Classic Method	3100	3100	2658	85.7
DS Approach	3100	3100	3014	97.2

Table 1: Comparison of T_{PR} for body classification, face classification, DS fusion and classic fusion

From Table 1, we can see that the DS fusion scheme gives an increase of approximately 11% in T_{PR} compared to the classic fusion scheme.

5 Conclusion

In this paper, we have proposed how to combine gender profiling classifier results by utilizing DS theory. We have used the cautious rule to combine gender profiling results from the same classifier at different time points and used Dempster's rule to combine profiling results from different classifiers. Experimental results show that the introduction of the DS theory indeed improves profiling performance.

We have mentioned that there are three problems that a classic gender profiling system should deal with, i.e., uncertain profiling results, unstable results over time for a gender profiling classifier, and different classifiers capturing different features. We have shown that a DS-based approach handles these three issues in a seamless way.

For future work, we plan to apply the fusion schemes to profiling classifier results generated from real video sequences.

Acknowledgement

This research work is sponsored by the EPSRC projects EP/G034303/1 and EP/H049606/1 (the CSIT project).

References

1. A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Statistics*, 28:325–339, 1967.

2. Thierry Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artifical Intelligence*, 172(2-3):234–264, 2008.
3. W. Liu, P. Miller, J. Ma, and W. Yan. Challenges of distributed intelligent surveillance system with heterogenous information. In *Procs. of QRASA*, pages 69–74, Pasadena, California, 2009.
4. J. Ma, W. Liu, and P. Miller. Event modelling and reasoning with uncertain information for distributed sensor networks. In *Procs. of SUM*, pages 236–249. Springer, 2010.
5. J. Ma, W. Liu, and P. Miller. Belief change with noisy sensing in the situation calculus. In *Procs. of UAI*, 2011.
6. J. Ma, W. Liu, and P. Miller. Handling sequential observations in intelligent surveillance. In *Proceedings of SUM*, pages 547–560, 2011.
7. J. Ma, W. Liu, P. Miller, and W. Yan. Event composition with imperfect information for bus surveillance. In *Procs. of AVSS*, pages 382–387. IEEE Press, 2009.
8. P. Miller, W. Liu, F. Fowler, H. Zhou, J. Shen, J. Ma, J. Zhang, W. Yan, K. McLaughlin, and S. Sezer. Intelligent sensor information system for public transport: To safely go .... In *Procs. of AVSS*, 2010.
9. ECIT Queen's University of Belfast. Airport corridor surveillance. *http://www.csit.qub.ac.uk/Research/ResearchGroups/IntelligentSurveillanceSystems*, 2010.
10. Gardiner Security. Glasgow transforms bus security with ip video surveillance. In *http://www.ipusergroup.com/doc-upload/Gardiner-Glasgowbuses.pdf*.
11. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
12. P. Smets. Non-standard logics for automated reasoning. In P. Smets, A. Mamdani, D. Dubois, and H. Prade, editors, *Belief functions*, pages 253–286, 1988.
13. H. Zhou, P. Miller, J. Zhang, M. Collins, and H. Wang. Gender classification using facial and full body features. In *Technical Report, CSIT, Queen's University Belfast, UK*, 2011.